ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Multivariate Analysis 94 (2005) 328–343
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Asymptotic robustness of the normal theory likelihood ratio statistic for two-level covariance structure models☆

Ke-Hai Yuan[a,*], Peter M. Bentler[b]

[a] *Department of Psychology, University of Notre Dame, Notre Dame, IN 46556-5645, USA*
[b] *University of California, Los Angeles, USA*

Received 5 August 2003

Abstract

Data in social and behavioral sciences are often hierarchically organized. Special statistical procedures have been developed to analyze such data while taking into account the resulting dependence of observations. Most of these developments require a multivariate normality distribution assumption. It is important to know whether normal theory-based inference can still be valid when applied to nonnormal hierarchical data sets. Using an analytical approach for balanced data and numerical illustrations for unbalanced data, this paper shows that the likelihood ratio statistic based on the normality assumption is asymptotically robust for many nonnormal distributions. The result extends the scope of asymptotic robustness theory that has been established in different contexts.

AMS 2000 subject classification: primary 62H10; secondary 62H15; 62H25

Keywords: Asymptotic robustness; Likelihood ratio statistic; Multilevel covariance structure; Nonnormal data

1. Introduction

Social science data often exhibit a hierarchical structure, and hence special models have been developed to analyze this kind of data [10,12,14,18,28]. Two related classes of such

☆ This research was supported by Grants DA01070 and DA00017 from the National Institute on Drug Abuse and a University of Notre Dame faculty research grant.

* Corresponding author. Fax: 574-631-8700.
E-mail address: kyuan@nd.edu (K.-H. Yuan)

0047-259X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2004.05.005

methods are the hierarchical linear model and the multilevel structural equation model [5,9,11,15,16,17,19,21,24–26] . All the above literature deals with model inference through methods that require a multivariate normality assumption for the hierarchical data. Typical data in practice seldom follow a multivariate normality. Although some procedures were developed for nonnormal data [27,37,39,40] statistical programs with the capability of handling multilevel data currently require the normality assumption for standard errors and test statistics. It is important to know whether results based on the normality assumption can still be valid when it is violated.

In the context of conventional linear latent variable models for independent observations [2], research on asymptotic robustness originated with [3,7]. In a factor analysis model, when the factors and errors are independent and errors are also independent among themselves, Amemiya and Anderson [1] found that the normal theory-based likelihood ratio (LR) statistic still asymptotically follows a chi-square distribution. This result was extended by Browne and Shapiro [8], Satorra and Bentler [32,33], Mooijaart and Bentler [22], Kano [13], Satorra [29,30,31], Satorra and Neudecker [34] and Yuan and Bentler [35] in the context of covariance structure analysis and by Yuan and Bentler [36,38] in the contexts of analyzing correlations and reliability coefficients. It is of interest to know whether a parallel result holds in the context of multilevel structural equation models (SEM). Simulation results in [37,39] indicate that the normal theory LR statistic still performs reasonably well when some independence condition holds. The purpose of this paper is to formally establish a set of conditions under which the normal theory LR statistics can be applied to a data set with heterogeneous skewnesses and kurtoses.

By introducing two classes of nonnormal distributions, Yuan and Bentler [35] gave very general characterizations of the asymptotic robustness of several statistics for conventional structural equation models. One class of their nonnormal distributions will be used here in studying the robustness of the LR statistic for multilevel SEM models. Yuan and Bentler [40] studied the behavior of the LR statistic within the class of elliptical distributions. When the between- and within-level components are not elliptically symmetric, the involved proof is more complicated. But the development is parallel to that in [40]. Specifically, we are only able to analytically establish the robustness property of the LR statistic when data are balanced. We resort to numerical illustrations when data are unbalanced.

2. The asymptotic distribution of the likelihood ratio statistic

Let the $p \times 1$ vectors $\mathbf{x}_{ij}, i = 1, \dots, n_j$ be observations from cluster j with $j = 1, \dots, J$. The 2-level structure of $\mathbf{x}_{ij}$ can be described by

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \mathbf{v}_j + \mathbf{u}_{ij}, \tag{1}$$

where $\boldsymbol{\mu}$ is a mean vector, $\mathbf{v}_j$ and $\mathbf{u}_{ij}$ are independent with $E(\mathbf{v}_j) = E(\mathbf{u}_{ij}) = \mathbf{0}$, $\text{Cov}(\mathbf{v}_j) = \boldsymbol{\Sigma}_b$ and $\text{Cov}(\mathbf{u}_{ij}) = \boldsymbol{\Sigma}_w$. Let $\boldsymbol{\theta}$ denote the vector of parameters in the structural model $\boldsymbol{\mu}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_b(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_w(\boldsymbol{\theta})$. The normal theory-based log likelihood function for cluster j was given in [37] as

$$l_j(\boldsymbol{\theta}) = c_j - \frac{1}{2} \log |\boldsymbol{\Sigma}_j(\boldsymbol{\theta})| - \frac{(n_j - 1)}{2} \log |\boldsymbol{\Sigma}_w(\boldsymbol{\theta})|$$

$$-\frac{1}{2}\sum_{i=1}^{n_j}(\mathbf{x}_{ij}-\bar{\mathbf{x}}_{.j})'\boldsymbol{\Sigma}_w^{-1}(\boldsymbol{\theta})(\mathbf{x}_{ij}-\bar{\mathbf{x}}_{.j})$$
$$-\frac{1}{2}[\bar{\mathbf{x}}_{.j}-\boldsymbol{\mu}(\boldsymbol{\theta})]'\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\theta})[\bar{\mathbf{x}}_{.j}-\boldsymbol{\mu}(\boldsymbol{\theta})],$$

where $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_b + n_j^{-1}\boldsymbol{\Sigma}_w$. The overall log likelihood function is

$$l(\boldsymbol{\theta}) = \sum_{j=1}^{J} l_j(\boldsymbol{\theta}).$$

For a $p \times p$ symmetric matrix $\mathbf{A}$, let vech($\mathbf{A}$) be the vector of stacking the columns of $\mathbf{A}$ that leaves the elements above the diagonal, and

$$\boldsymbol{\sigma}_j = \mathrm{vech}(\boldsymbol{\Sigma}_j), \quad \boldsymbol{\sigma}_b = \mathrm{vech}(\boldsymbol{\Sigma}_b), \quad \boldsymbol{\sigma}_w = \mathrm{vech}(\boldsymbol{\Sigma}_w).$$

When the model is saturated, the parameter vector is $\boldsymbol{\beta} = (\boldsymbol{\mu}', \boldsymbol{\sigma}_b', \boldsymbol{\sigma}_w')'$. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimates (MLE) of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, respectively. Yuan and Bentler [37] expressed the likelihood ratio statistic $T_{\mathrm{LR}} = 2[l(\hat{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\theta}})]$ as a quadratic form in $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. We need some notation before introducing this quadratic form.

Notice that $l(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ defined through $\boldsymbol{\beta}(\boldsymbol{\theta}) = (\boldsymbol{\mu}'(\boldsymbol{\theta}), \boldsymbol{\sigma}_b'(\boldsymbol{\theta}), \boldsymbol{\sigma}_w'(\boldsymbol{\theta}))'$. We will use a dot on top of a function to denote the derivative as in $\dot{\boldsymbol{\beta}}(\boldsymbol{\theta}) = d\boldsymbol{\beta}(\boldsymbol{\theta})/d\boldsymbol{\theta}$. The population values of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ corresponding to correctly specified models are denoted by $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}_0$, respectively. Thus, $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$. When a function is evaluated at the population value, we will omit its argument, for example, $\dot{\boldsymbol{\beta}} = \dot{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$. Let vec($\mathbf{A}$) be the vector formed by stacking the columns of $\mathbf{A}$. Then there is a duplication matrix $\mathbf{D}_p$ such that $\mathrm{vec}(\mathbf{A}) = \mathbf{D}_p\,\mathrm{vech}(\mathbf{A})$ [20]. Denote

$$\mathbf{W}_j = 2^{-1}\mathbf{D}_p'(\boldsymbol{\Sigma}_j^{-1}\otimes\boldsymbol{\Sigma}_j^{-1})\mathbf{D}_p, \quad \mathbf{W}_w = 2^{-1}\mathbf{D}_p'(\boldsymbol{\Sigma}_w^{-1}\otimes\boldsymbol{\Sigma}_w^{-1})\mathbf{D}_p.$$

Then the likelihood ratio statistic can be expressed as [37]

$$T_{\mathrm{LR}} = \sqrt{J}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{U}_J\sqrt{J}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1), \tag{2a}$$

where $o_p(1)$ is a term that approaches zero in probability as $J \to \infty$ (see e.g. Chapter 14 of [6]),

$$\mathbf{U}_J = \mathbf{A}_J - \mathbf{A}_J\dot{\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}'\mathbf{A}_J\dot{\boldsymbol{\beta}})^{-1}\dot{\boldsymbol{\beta}}'\mathbf{A}_J \tag{2b}$$

with

$$\mathbf{A}_J = \begin{pmatrix} \mathbf{A}_{J11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{J22} & \mathbf{A}_{J23} \\ \mathbf{0} & \mathbf{A}_{J32} & \mathbf{A}_{J33} \end{pmatrix} \tag{2c}$$

and

$$\mathbf{A}_{J11} = J^{-1}\sum_{j=1}^{J}\boldsymbol{\Sigma}_j^{-1}, \quad \mathbf{A}_{J22} = J^{-1}\sum_{j=1}^{J}\mathbf{W}_j,$$

$$\mathbf{A}_{J23} = J^{-1}\sum_{j=1}^{J} n_j^{-1}\mathbf{W}_j, \quad \mathbf{A}_{J33} = J^{-1}\sum_{j=1}^{J}[n_j^{-2}\mathbf{W}_j + (n_j-1)\mathbf{W}_w].$$

Let $\mathbf{g}_j = (\mathbf{g}'_{j1}, \mathbf{g}'_{j2}, \mathbf{g}'_{j3})'$ with

$$\mathbf{g}_{j1} = \boldsymbol{\Sigma}_j^{-1}(\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu}),$$

$$\mathbf{g}_{j2} = \mathbf{W}_j \,\mathrm{vech}\{(\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu})(\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu})' - \boldsymbol{\Sigma}_j\},$$

$$\mathbf{g}_{j3} = n_j^{-1}\mathbf{W}_j \,\mathrm{vech}\{(\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu})(\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu})' - \boldsymbol{\Sigma}_j\} + \mathbf{W}_w \,\mathrm{vech}\{n_j\mathbf{S}_j - (n_j - 1)\boldsymbol{\Sigma}_w\},$$

where $\mathbf{S}_j = n_j^{-1} \sum_{i=1}^{n_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{.j})(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{.j})'$, and

$$\mathbf{G}_J = J^{-1} \sum_{j=1}^{J} \mathbf{g}_j. \tag{3}$$

Then the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is given by (see [37])

$$\sqrt{J}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Gamma}), \tag{4}$$

where $\boldsymbol{\Gamma} = \lim_{J\to\infty} \mathbf{A}_J^{-1}\mathbf{B}_J\mathbf{A}_J^{-1}$ with $\mathbf{B}_J = J\,\mathrm{Cov}(\mathbf{G}_J)$. Notice that the form of $\mathbf{A}_J$ does not depend on the specific distribution of $\mathbf{x}_{ij}$, neither does the matrix $\dot{\boldsymbol{\beta}}$ in (2b). It follows from (2a) and (2b) that the asymptotic distribution of T_{LR} depends on the distribution of $\mathbf{x}_{ij}$ only through $\hat{\boldsymbol{\beta}}$, whose asymptotic distribution is characterized by $\boldsymbol{\Gamma}$ as in (4). Thus, the asymptotic distribution of T_{LR} depends on the distribution of $\mathbf{x}_{ij}$ only through $\mathbf{B}_J$. Notice that $\mathbf{B}_J$ is of dimension $[p(p+2)] \times [p(p+2)]$. When $J > p(p+2)$, $\mathbf{B}_J$ is of full rank with probability one and so is $\boldsymbol{\Gamma}_J = \mathbf{A}_J^{-1}\mathbf{B}_J\mathbf{A}_J^{-1}$. Let $\mathbf{z}_J = \boldsymbol{\Gamma}_J^{-1/2}\sqrt{J}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, then $\mathbf{z}_J \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{p(p+2)})$ and

$$T_{\mathrm{LR}} = \mathbf{z}'_J\boldsymbol{\Gamma}_J^{1/2}\mathbf{U}_J\boldsymbol{\Gamma}_J^{1/2}\mathbf{z}_J + o_p(1), \tag{5}$$

where $\mathbf{I}_{p(p+2)}$ is the identity matrix of order $p(p+2)$. The right-hand side of (5) is a quadratic form in $\mathbf{z}_J$, so the asymptotic distribution of T_{LR} is decided by the eigenvalues of $\mathbf{U}\boldsymbol{\Gamma} = \lim_{J\to\infty} \mathbf{U}_J\boldsymbol{\Gamma}_J$ (see Section 1.4 of [23]). Asymptotic robustness holds only when all the nonzero eigenvalues of $\mathbf{U}\boldsymbol{\Gamma}$ equal 1.0.

Parallel to conventional covariance structure models, we assume $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$ is saturated and parameters $\boldsymbol{\theta}_b$ and $\boldsymbol{\theta}_w$ in $\boldsymbol{\Sigma}_b(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_b(\boldsymbol{\theta}_b)$ and $\boldsymbol{\Sigma}_w(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_w(\boldsymbol{\theta}_w)$ are functionally independent. So $\boldsymbol{\theta} = (\boldsymbol{\mu}', \boldsymbol{\theta}'_b, \boldsymbol{\theta}'_w)'$ and

$$\dot{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dot{\boldsymbol{\sigma}}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dot{\boldsymbol{\sigma}}_w \end{pmatrix}.$$

Let q_b and q_w be the number of parameters in $\boldsymbol{\theta}_b$ and $\boldsymbol{\theta}_w$. Then $\dot{\boldsymbol{\beta}}$ is a matrix of order $[p(p+2)] \times (p + q_b + q_w)$.

In order to characterize the distribution of T_{LR} we will introduce a class of nonnormal distributions given by Yuan and Bentler [35]. The asymptotic robustness of T_{LR} will be studied within this class of distributions.

Let $\xi_1, \ldots, \xi_m$ be independent random variables with $E(\xi_i) = 0$, $E(\xi_i^2) = 1$, $E(\xi_i^3) = \zeta_i$, $E(\xi_i^4) = \kappa_i$, and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)'$. Let r be a random variable which is independent of $\boldsymbol{\xi}$, $E(r^2) = 1$, $E(r^3) = \gamma$, and $E(r^4) = \tau$. Also, let $m \geqslant p$ and $\mathbf{T} = (t_{ij}) = (\mathbf{t}_1, \ldots, \mathbf{t}_m)$ be a $p \times m$ matrix of rank p such that $\mathbf{TT}' = \boldsymbol{\Sigma}$, where $\mathbf{t}_j = (t_{1j}, \ldots, t_{pj})'$. Then the random vector

$$\mathbf{x} = r\mathbf{T}\boldsymbol{\xi} \tag{6}$$

will generally follow a nonnormal distribution. Different distributions will be obtained by choosing a different set of ξ_i's, $\mathbf{T}$ and r. It is easy to see that the population covariance matrix of $\mathbf{x}$ is given by $\boldsymbol{\Sigma}$. Yuan and Bentler [35] obtained the fourth-order moment matrix $\boldsymbol{\Omega} = \mathrm{Cov}[\mathrm{vech}(\mathbf{xx}')]$ as

$$\boldsymbol{\Omega} = 2\tau\mathbf{D}_p^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}_p^{+'} + (\tau - 1)\boldsymbol{\sigma\sigma}' + \tau \sum_{i=1}^{m} (\kappa_i - 3)\mathrm{vech}(\mathbf{t}_i\mathbf{t}_i')\mathrm{vech}'(\mathbf{t}_i\mathbf{t}_i'). \tag{7}$$

We assume that the between-level vector $\mathbf{v}_j$ follows (6) and has a fourth-order moment matrix

$$\begin{aligned}\boldsymbol{\Omega}_b = {} & 2\tau\mathbf{D}_p^+(\boldsymbol{\Sigma}_b \otimes \boldsymbol{\Sigma}_b)\mathbf{D}_p^{+'} + (\tau - 1)\boldsymbol{\sigma}_b\boldsymbol{\sigma}_b' \\ & + \tau \sum_{i=1}^{m_b} (\kappa_i^{(b)} - 3)\mathrm{vech}(\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime})\mathrm{vech}'(\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime}),\end{aligned} \tag{8}$$

and the within-level vector $\mathbf{u}_{ij}$ also follows (6) and has a fourth-order moment matrix

$$\begin{aligned}\boldsymbol{\Omega}_w = {} & 2\tau\mathbf{D}_p^+(\boldsymbol{\Sigma}_w \otimes \boldsymbol{\Sigma}_w)\mathbf{D}_p^{+'} + (\tau - 1)\boldsymbol{\sigma}_w\boldsymbol{\sigma}_w' \\ & + \tau \sum_{i=1}^{m_w} (\kappa_i^{(w)} - 3)\mathrm{vech}(\mathbf{t}_i^{(w)}\mathbf{t}_i^{(w)\prime})\mathrm{vech}'(\mathbf{t}_i^{(w)}\mathbf{t}_i^{(w)\prime}).\end{aligned} \tag{9}$$

Notice that we have assumed the r_b that generates $\mathbf{v}_j$ has the same fourth-order moment as the r_w that generates $\mathbf{u}_{ij}$, but they can have different third-order moments. The matrices $\mathbf{T}_b = (t_{ij}^{(b)})$ and $\mathbf{T}_w = (t_{ij}^{(w)})$ can be of different dimensions although they need to satisfy $\mathbf{T}_b\mathbf{T}_b{}' = \boldsymbol{\Sigma}_b$ and $\mathbf{T}_w\mathbf{T}_w{}' = \boldsymbol{\Sigma}_w$. The corresponding $\boldsymbol{\xi}_b$ and $\boldsymbol{\xi}_w$ can be totally different random vectors.

We need to obtain the matrix $\mathbf{B}_J$ when $\mathrm{Cov}[\mathrm{vech}(\mathbf{v}_j\mathbf{v}_j')] = \boldsymbol{\Omega}_b$ and $\mathrm{Cov}[\mathrm{vech}(\mathbf{u}_{ij}\mathbf{u}_{ij}')] = \boldsymbol{\Omega}_w$. Note that the $\mathbf{G}_J$ in (3) involves the random terms $\mathbf{h}_{j1} = (\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu})$, $\mathbf{h}_{j2} = \mathrm{vech}[(\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu})(\bar{\mathbf{x}}_{.j} - \boldsymbol{\mu})' - \boldsymbol{\Sigma}_j]$, $\mathbf{h}_{j3} = \mathrm{vech}[n_j\mathbf{S}_j - (n_j - 1)\boldsymbol{\Sigma}_w]$. Expressions for $\mathbf{C}_{jkl} = E(\mathbf{h}_{jk}\mathbf{h}_{jl}')$ are given in the following lemma.

Lemma 1. *Let the* $\mathbf{v}_j$ *and* $\mathbf{u}_{ij}$ *follow the distributions generated by* (6) *whose fourth-order moment matrices are, respectively, given by* (8) *and* (9). *Then* $\mathbf{C}_{j11} = \boldsymbol{\Sigma}_j$,

$$\mathbf{C}_{j12} = \gamma_b\mathbf{T}_b\boldsymbol{\Delta}_b(\mathbf{T}_b' \otimes \mathbf{T}_b')\mathbf{D}_p^{+'} + \frac{1}{n_j^2}\gamma_w\mathbf{T}_w\boldsymbol{\Delta}_w(\mathbf{T}_w' \otimes \mathbf{T}_w')\mathbf{D}_p^{+'},$$

where $\boldsymbol{\Delta}_b = (\delta_{ij}^{(b)})$ *is the* $m_b \times [m_b(m_b + 1)/2]$ *matrix such that*

$$\delta_{ij}^{(b)} = \begin{cases} \zeta_i^{(b)} & j = (i-1)m_b + i, \\ 0 & elsewhere \end{cases}$$

and $\boldsymbol{\Delta}_w = (\delta_{ij}^{(w)})$ *is similarly defined,*

$$\mathbf{C}_{j13} = \left(1 - \frac{1}{n_j}\right)\gamma_w \mathbf{T}_w \boldsymbol{\Delta}_w (\mathbf{T}_w' \otimes \mathbf{T}_w')\mathbf{D}_p^{+'},$$

$$\begin{aligned}\mathbf{C}_{j22} = \boldsymbol{\Omega}_b &+ \frac{1}{n_j^3}\,\boldsymbol{\Omega}_w + \frac{2(n_j-1)}{n_j^3}\,\mathbf{D}_p^+(\boldsymbol{\Sigma}_w \otimes \boldsymbol{\Sigma}_w)\mathbf{D}_p^{+'}\\ &+ \frac{2}{n_j}\,\mathbf{D}_p^+[(\boldsymbol{\Sigma}_w \otimes \boldsymbol{\Sigma}_b) + (\boldsymbol{\Sigma}_b \otimes \boldsymbol{\Sigma}_w)]\mathbf{D}_p^{+'},\end{aligned}$$

$$\mathbf{C}_{j23} = \frac{(n_j-1)}{n_j^2}\,\boldsymbol{\Omega}_w - \frac{2(n_j-1)}{n_j^2}\,\mathbf{D}_p^+(\boldsymbol{\Sigma}_w \otimes \boldsymbol{\Sigma}_w)\mathbf{D}_p^{+'},$$

$$\mathbf{C}_{j33} = \left(n_j - 2 + \frac{1}{n_j}\right)\boldsymbol{\Omega}_w + 2\left(1 - \frac{1}{n_j}\right)\mathbf{D}_p^+(\boldsymbol{\Sigma}_w \otimes \boldsymbol{\Sigma}_w)\mathbf{D}_p^{+'}.$$

In the appendix we give an outline for obtaining $\mathbf{C}_{j12}$ and $\mathbf{C}_{j13}$. Using (8) and (9) for the fourth-order moments of $\mathbf{v}_j$ and $\mathbf{u}_{ij}$, $\mathbf{C}_{j11}$, $\mathbf{C}_{j22}$, $\mathbf{C}_{j23}$ and $\mathbf{C}_{j33}$ can be obtained directly following from the appendix of Yuan and Bentler [40]. Because the specific forms of $\mathbf{C}_{j12}$ and $\mathbf{C}_{j13}$ are not used in characterizing the asymptotic distribution of T_{LR}, we will denote them by $\mathbf{C}_{j12} = \mathbf{O}(1)$ and $\mathbf{C}_{j13} = \mathbf{O}(1)$ to imply that all their elements are bounded.

Let

$$\mathbf{B}_J = \begin{pmatrix} \mathbf{B}_{J11} & \mathbf{B}_{J12} & \mathbf{B}_{J13} \\ \mathbf{B}_{J21} & \mathbf{B}_{J22} & \mathbf{B}_{J23} \\ \mathbf{B}_{J31} & \mathbf{B}_{J32} & \mathbf{B}_{J33} \end{pmatrix}, \tag{10a}$$

it follows from Lemma 1 that

$$\mathbf{B}_{J11} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\Sigma}_j^{-1}, \tag{10b}$$

$$\mathbf{B}_{J22} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{W}_j \mathbf{C}_{j22} \mathbf{W}_j, \tag{10c}$$

$$\mathbf{B}_{J23} = \frac{1}{J}\sum_{j=1}^{J} n_j^{-1}\mathbf{W}_j \mathbf{C}_{j22} \mathbf{W}_j + \frac{1}{J}\sum_{j=1}^{J} \mathbf{W}_j \mathbf{C}_{j23} \mathbf{W}_w, \tag{10d}$$

$$\mathbf{B}_{J33} = \frac{1}{J}\sum_{j=1}^{J} (n_j^{-2}\mathbf{W}_j \mathbf{C}_{j22} \mathbf{W}_j + n_j^{-1}\mathbf{W}_j \mathbf{C}_{j23} \mathbf{W}_w + n_j^{-1}\mathbf{W}_w \mathbf{C}_{j32} \mathbf{W}_j + \mathbf{W}_w \mathbf{C}_{j33} \mathbf{W}_w). \tag{10e}$$

Similarly, the specific form of $\mathbf{B}_{J12}$ and $\mathbf{B}_{J13}$ are not needed in studying the distribution of T_{LR}. They can be denoted by $\mathbf{B}_{J12} = \mathbf{O}(1)$ and $\mathbf{B}_{J13} = \mathbf{O}(1)$.

Yuan and Bentler [37] argued that one needs to have a large J in order for $\hat{\boldsymbol{\theta}}$ to be in the neighborhood of $\boldsymbol{\theta}_0$ or the distribution of $\hat{\boldsymbol{\theta}}$ to be approximately normal. A large J also guarantees that the matrix $\mathbf{U}\boldsymbol{\Gamma}$ is well-approximated by $\mathbf{U}_J\boldsymbol{\Gamma}_J$. In the following, we will study the eigenvalues of $\mathbf{U}\boldsymbol{\Gamma}$ via those of $\mathbf{U}_J\boldsymbol{\Gamma}_J$, within the class of distributions specified in (6).

When n_j's are not equal, the covariance matrix $\boldsymbol{\Gamma}_J = \mathbf{A}_J^{-1}\mathbf{B}_J\mathbf{A}_J^{-1}$ is too complicated to work with analytically. Hence, in the rest of this section, we only deal with the simple case of balanced data with $n_j = n$. For balanced data, $\boldsymbol{\Sigma}_j$, $\mathbf{W}_j$, $\mathbf{C}_{j22}$, $\mathbf{C}_{j23}$, $\mathbf{C}_{j32}$, $\mathbf{C}_{j33}$ do not depend on j. They are functions of n instead. It follows from (2) and (10) that the index J is not involved in the expressions of $\mathbf{A}_J$, $\mathbf{B}_J$ or $\mathbf{U}_J$. So we will index them by $\mathbf{A}_n$, $\mathbf{B}_n$ and $\mathbf{U}_n$, and denote $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_j$ and $\mathbf{W}_n = \mathbf{W}_j$. The following analysis will show that, as $n \to \infty$, $\mathbf{U}_n\boldsymbol{\Gamma}_n$ converges to a matrix whose nonzero eigenvalues can be equal. Notice that the convergence in (4) is stochastic while the convergence of $\mathbf{U}_n\boldsymbol{\Gamma}_n$ is nonstochastic.

Let

$$\mathbf{M}_n = \mathbf{I}_{p(p+2)} - \mathbf{A}_n\dot{\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}'\mathbf{A}_n\dot{\boldsymbol{\beta}})^{-1}\dot{\boldsymbol{\beta}}'.$$

Yuan and Bentler [40] showed that

$$\mathbf{M}_n = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{L}_w \end{pmatrix} + \mathbf{o}(1), \tag{11}$$

where

$$\mathbf{L}_b = \mathbf{I}_{p^*} - \mathbf{W}_b\dot{\boldsymbol{\sigma}}_b(\dot{\boldsymbol{\sigma}}_b'\mathbf{W}_b\dot{\boldsymbol{\sigma}}_b)^{-1}\dot{\boldsymbol{\sigma}}_b',$$

$$\mathbf{L}_w = \mathbf{I}_{p^*} - \mathbf{W}_w\dot{\boldsymbol{\sigma}}_w(\dot{\boldsymbol{\sigma}}_w'\mathbf{W}_w\dot{\boldsymbol{\sigma}}_w)^{-1}\dot{\boldsymbol{\sigma}}_w'$$

with $p^* = p(p+1)/2$, and $\mathbf{o}(1)$ denotes a matrix whose elements all approach zero when $n \to \infty$. It is easy to see that

$$\mathbf{U}_n\boldsymbol{\Gamma}_n = \mathbf{M}_n(\mathbf{A}_n\boldsymbol{\Gamma}_n) = \mathbf{M}_n(\mathbf{B}_n\mathbf{A}_n^{-1}). \tag{12}$$

We need to find the limit of $\mathbf{F}_n = \mathbf{B}_n\mathbf{A}_n^{-1}$ before obtaining the limit of $\mathbf{U}_n\boldsymbol{\Gamma}_n$.

Using the rule of matrix inversion for partitioned matrices [20, p. 11], we have

$$\mathbf{A}_n^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_n^{-1} + [n^2(n-1)]^{-1}\mathbf{W}_w^{-1} & -[n(n-1)]^{-1}\mathbf{W}_w^{-1} \\ \mathbf{0} & -[n(n-1)]^{-1}\mathbf{W}_w^{-1} & (n-1)^{-1}\mathbf{W}_w^{-1} \end{pmatrix}.$$

Let

$$\mathbf{F}_n = \begin{pmatrix} \mathbf{I}_p & \mathbf{F}_{n12} & \mathbf{F}_{n13} \\ \mathbf{F}_{n21} & \mathbf{F}_{n22} & \mathbf{F}_{n23} \\ \mathbf{F}_{n31} & \mathbf{F}_{n32} & \mathbf{F}_{n33} \end{pmatrix}.$$

Following the rule of matrix multiplication we obtain

$$\mathbf{F}_{n22} = \mathbf{W}_n\mathbf{C}_{n22} - \frac{1}{n(n-1)}\mathbf{W}_n\mathbf{C}_{n23}, \tag{13}$$

$$\mathbf{F}_{n23} = \frac{1}{(n-1)}\mathbf{W}_n\mathbf{C}_{n23}, \tag{14}$$

$$\mathbf{F}_{n32} = \frac{1}{n}\mathbf{W}_n\mathbf{C}_{n22} + \mathbf{W}_n\mathbf{C}_{n32} - \frac{1}{n^2(n-1)}\mathbf{W}_n\mathbf{C}_{n23} - \frac{1}{n(n-1)}\mathbf{W}_n\mathbf{C}_{n33}, \tag{15}$$

$$\mathbf{F}_{n33} = \frac{1}{n(n-1)}\mathbf{W}_n\mathbf{C}_{n23} + \frac{1}{n-1}\mathbf{W}_w\mathbf{C}_{n33}. \tag{16}$$

The specific forms of $\mathbf{F}_{n12}$, $\mathbf{F}_{n13}$, $\mathbf{F}_{n21}$ and $\mathbf{F}_{n31}$ are not needed and they can be denoted by $\mathbf{F}_{n12} = \mathbf{O}(1)$, $\mathbf{F}_{n13} = \mathbf{O}(n^{-1})$, $\mathbf{F}_{n21} = \mathbf{O}(1)$ and $\mathbf{F}_{n31} = \mathbf{O}(1)$.

We need to calculate the following matrices in order to further simplify the form of $\mathbf{F}_n$. Notice that $\mathbf{W}_n^{-1} = 2\mathbf{D}_p^+(\boldsymbol{\Sigma}_n \otimes \boldsymbol{\Sigma}_n)\mathbf{D}_p^{+'}$ and $\mathbf{W}_n = \mathbf{W}_b + \mathbf{O}(n^{-1})$. It follows from Lemma 1 that

$$\mathbf{W}_n\mathbf{C}_{n22} = \tau\mathbf{I}_{p^*} + \mathbf{H}_b + \mathbf{O}\left(\frac{1}{n}\right), \tag{17}$$

$$\mathbf{W}_n\mathbf{C}_{n23} = \mathbf{O}\left(\frac{1}{n}\right), \quad \mathbf{W}_n\mathbf{C}_{n32} = \mathbf{O}\left(\frac{1}{n}\right), \quad \mathbf{W}_n\mathbf{C}_{n33} = \mathbf{O}(n), \tag{18}$$

$$\mathbf{W}_w\mathbf{C}_{n33} = n\left(1-\frac{1}{n}\right)^2(\tau\mathbf{I}_{p^*} + \mathbf{H}_w) + \left(1-\frac{1}{n}\right)\mathbf{I}_{p^*}, \tag{19}$$

where

$$\mathbf{H}_b = (\tau-1)\mathbf{W}_b\boldsymbol{\sigma}_b\boldsymbol{\sigma}_b' + \tau\sum_{i=1}^{m_b}(\kappa_i^{(b)} - 3)\mathbf{W}_b\,\mathrm{vech}(\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime})\mathrm{vech}'(\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime})$$

and

$$\mathbf{H}_w = (\tau-1)\mathbf{W}_w\boldsymbol{\sigma}_w\boldsymbol{\sigma}_w' + \tau\sum_{i=1}^{m_w}(\kappa_i^{(w)} - 3)\mathbf{W}_w\,\mathrm{vech}(\mathbf{t}_i^{(w)}\mathbf{t}_i^{(w)\prime})\mathrm{vech}'(\mathbf{t}_i^{(w)}\mathbf{t}_i^{(w)\prime}).$$

It follows from (13), (17) and (18) that

$$\mathbf{F}_{n22} = \tau\mathbf{I}_{p^*} + \mathbf{H}_b + \mathbf{O}\left(\frac{1}{n}\right). \tag{20}$$

It follows from (14) and (18) that

$$\mathbf{F}_{n23} = \mathbf{O}\left(\frac{1}{n^2}\right),$$

and from (15), (17) and (18) that

$$\mathbf{F}_{n32} = \mathbf{O}\left(\frac{1}{n}\right).$$

It follows from (16), (18) and (19) that

$$\mathbf{F}_{n33} = \tau\mathbf{I}_{p^*} + \mathbf{H}_w + \mathbf{O}\left(\frac{1}{n}\right). \tag{21}$$

It follows from (20) and (21) that

$$\mathbf{F}_n = \begin{pmatrix} \mathbf{I}_p & \mathbf{O}(1) & \mathbf{O}(n^{-1}) \\ \mathbf{O}(1) & \tau\mathbf{I}_{p^*} + \mathbf{H}_b + \mathbf{O}(n^{-1}) & \mathbf{O}(n^{-2}) \\ \mathbf{O}(1) & \mathbf{O}(n^{-1}) & \tau\mathbf{I}_{p^*} + \mathbf{H}_w + \mathbf{O}(n^{-1}) \end{pmatrix}. \tag{22}$$

Let

$$\mathbf{V}_b = \mathbf{W}_b - \mathbf{W}_b\dot{\boldsymbol{\sigma}}_b(\dot{\boldsymbol{\sigma}}_b'\mathbf{W}_b\dot{\boldsymbol{\sigma}}_b)^{-1}\dot{\boldsymbol{\sigma}}_b'\mathbf{W}_b$$

and

$$\mathbf{V}_w = \mathbf{W}_w - \mathbf{W}_w\dot{\boldsymbol{\sigma}}_w(\dot{\boldsymbol{\sigma}}_w'\mathbf{W}_w\dot{\boldsymbol{\sigma}}_w)^{-1}\dot{\boldsymbol{\sigma}}_w'\mathbf{W}_w.$$

We need to assume

$$\mathbf{V}_b\,\mathrm{vech}(\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime}) = \mathbf{0},\ i = 1,\ldots,m_b; \quad \mathbf{V}_w\,\mathrm{vech}(\mathbf{t}_j^{(w)}\mathbf{t}_j^{(w)\prime}) = \mathbf{0},\ j = 1,\ldots,m_w. \tag{23}$$

Because

$$\boldsymbol{\Sigma}_b = \sum_{i=1}^{m_b}\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime} \quad \text{and} \quad \boldsymbol{\Sigma}_w = \sum_{i=1}^{m_w}\mathbf{t}_i^{(w)}\mathbf{t}_i^{(w)\prime},$$

conditions in (23) also imply $\mathbf{V}_b\boldsymbol{\sigma}_b = \mathbf{0}$ and $\mathbf{V}_w\boldsymbol{\sigma}_w = \mathbf{0}$.

Combining (11), (12) and (22) yields

$$\mathbf{U}_n\boldsymbol{\Gamma}_n = \mathbf{M}_n\mathbf{F}_n = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{O}(1) & \tau\mathbf{L}_b + \mathbf{L}_b\mathbf{H}_b & \mathbf{0} \\ \mathbf{O}(1) & \mathbf{0} & \tau\mathbf{L}_w + \mathbf{L}_w\mathbf{H}_w \end{pmatrix} + \mathbf{o}(1). \tag{24}$$

Conditions in (23) lead to

$$\mathbf{L}_b\mathbf{H}_b = \mathbf{0} \quad \text{and} \quad \mathbf{L}_w\mathbf{H}_w = \mathbf{0}.$$

Rewrite the first term on the right-hand side of (24) as

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_b^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}_w^{1/2} \end{pmatrix}\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_w \end{pmatrix}\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W}_b^{-1/2}\mathbf{O}(1) & \tau\mathbf{W}_b^{-1/2} & \mathbf{0} \\ \mathbf{W}_w^{-1/2}\mathbf{O}(1) & \mathbf{0} & \tau\mathbf{W}_w^{-1/2} \end{pmatrix}, \tag{25}$$

where

$$\mathbf{Q}_b = \mathbf{I}_{p^*} - \mathbf{W}_b^{1/2}\dot{\boldsymbol{\sigma}}_b(\dot{\boldsymbol{\sigma}}_b'\mathbf{W}_b\dot{\boldsymbol{\sigma}}_b)^{-1}\dot{\boldsymbol{\sigma}}_b'\mathbf{W}_b^{1/2}$$

and

$$\mathbf{Q}_w = \mathbf{I}_{p^*} - \mathbf{W}_w^{1/2}\dot{\boldsymbol{\sigma}}_w(\dot{\boldsymbol{\sigma}}_w'\mathbf{W}_w\dot{\boldsymbol{\sigma}}_w)^{-1}\dot{\boldsymbol{\sigma}}_w'\mathbf{W}_w^{1/2}$$

are projection matrices. Because the eigenvalues of $\mathbf{XY}$ equal the eigenvalues of $\mathbf{YX}$, it follows from (25) that the eigenvalues of $\lim_{n\to\infty}\mathbf{U}_n\boldsymbol{\Gamma}_n$ are identical to those of

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau\mathbf{Q}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tau\mathbf{Q}_q \end{pmatrix}.$$

As a projection matrix, $\mathbf{Q}_b$ has $p^* - q_b$ nonzero eigenvalues of 1.0, and similarly $\mathbf{Q}_w$ has $p^* - q_w$ nonzero eigenvalues of 1.0, thus $\lim_{n\to\infty}\mathbf{U}_n\boldsymbol{\Gamma}_n$ has $p(p+1) - (q_b + q_w)$ nonzero eigenvalues of τ. This leads to the following theorem for balanced data:

Theorem 1. *Let* $\mathbf{v}_j$ *and* $\mathbf{u}_{ij}$*, respectively, follow distributions generated by* (6) *with* $E(r_b^4) = E(r_w^4) = \tau$. *When* $\boldsymbol{\theta}_b$ *and* $\boldsymbol{\theta}_w$ *are functionally independent and conditions in* (23) *are satisfied, as* $n \to \infty$ *and* $J \to \infty$*, the statistic* T_{LR} *approaches* $\tau\chi^2_{p(p+1)-(q_b+q_w)}$.

The theorem immediately implies the following corollary:

Corollary 1. *Let* $\mathbf{v}_j$ *and* $\mathbf{u}_{ij}$, *respectively, follow distributions generated by* (6) *with* $r_b = r_w = 1.0$. *When* $\boldsymbol{\theta}_b$ *and* $\boldsymbol{\theta}_w$ *are functionally independent and conditions in* (23) *are satisfied, then the statistic* T_{LR} *is asymptotically robust.*

The condition $r_b = r_w = 1.0$ in the above corollary can also be replaced by $E(r_b^4) = E(r_w^4) = 1.0$. Notice that $E(r_b^2) = E(r_w^2) = 1.0$. This implies $\mathrm{Var}(r_b^2) = \mathrm{Var}(r_w^2) = 1.0$. Thus $\Pr(r_b = \pm 1) = \Pr(r_w = \pm 1) = 1.0$, which is slightly more general than $r_b = r_w = 1.0$.

We may wonder how many distributions in (6) will satisfy the conditions specified in (23). For the given model $\boldsymbol{\Sigma}_b(\boldsymbol{\theta})$, $\mathbf{V}_b$ is fixed. Conditions in (23) put m_b constraints on the coefficients $t_{ij}^{(b)}$. The condition $\mathbf{T}_b\mathbf{T}_b{}' = \boldsymbol{\Sigma}_b$ puts another $p(p+1)/2$ constraints on the $t_{ij}^{(b)}$. So the pm_b coefficients in $\mathbf{T}_b$ need to satisfy $m_b + p(p+1)/2$ equations. Thus, the solutions to $t_{ij}^{(b)}$ occupy a space of dimension $(p-1)m_b - p(p+1)/2$. Since one can arbitrarily choose m_b as long as it is no less than p, there are infinitely many $t_{ij}^{(b)}$ to choose. Similarly, there are infinitely many $t_{ij}^{(w)}$ to choose. For given $\mathbf{T}_b$ and $\mathbf{T}_w$, one can arbitrarily choose $\boldsymbol{\xi}_b$ and $\boldsymbol{\xi}_w$ as long as $\kappa_i^{(b)}$ and $\kappa_i^{(w)}$ are bounded. So the LR statistic can be asymptotically valid for many nonnormal distributions.

This theory can be illustrated with a two-level confirmatory factor model

$$\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}_b\boldsymbol{\Phi}_b\boldsymbol{\Lambda}_b' + \boldsymbol{\Psi}_b \quad \text{and} \quad \boldsymbol{\Sigma}_w = \boldsymbol{\Lambda}_w\boldsymbol{\Phi}_w\boldsymbol{\Lambda}_w' + \boldsymbol{\Psi}_w, \tag{26}$$

where $\boldsymbol{\Lambda}_b$ and $\boldsymbol{\Lambda}_w$ are factor loading matrices, $\boldsymbol{\Phi}_b$ and $\boldsymbol{\Phi}_w$ are factor covariance matrices; $\boldsymbol{\Psi}_b$ and $\boldsymbol{\Psi}_w$ are the unique variance matrices. Suppose there are k factors at the between-level. A popular structure for $\boldsymbol{\Lambda}_b$ is

$$\boldsymbol{\Lambda}_b = \begin{pmatrix} \boldsymbol{\lambda}_1^{(b)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{\lambda}_k^{(b)} \end{pmatrix},$$

where $\boldsymbol{\lambda}_j^{(b)} = (\lambda_{1j}^{(b)}, \ldots, \lambda_{k_j j}^{(b)})'$. That is, each variable in $\mathbf{v}_j$ only depends on one common factor. In order for model $\boldsymbol{\Sigma}_b$ in (26) to be identifiable, it is necessary to fix the scale of each factor. This can be obtained by fixing at 1.0 the last element $\lambda_{k_j j}^{(b)}$ in each $\boldsymbol{\lambda}_j^{(b)}$. Under this setup, the partial derivative of $\boldsymbol{\Sigma}_b$ with respect to $\lambda_{ij}^{(b)}$ is given by

$$\dot{\boldsymbol{\Sigma}}_{b\lambda_{ij}} = (\mathbf{0}_{p\times(j-1)}, \mathbf{e}_s, \mathbf{0}_{p\times(k-j)})\boldsymbol{\Phi}_b\boldsymbol{\Lambda}_b' + \boldsymbol{\Lambda}_b\boldsymbol{\Phi}_b(\mathbf{0}_{p\times(j-1)}, \mathbf{e}_s, \mathbf{0}_{p\times(k-j)})' \tag{27a}$$

for $i = 1, \ldots, k_j - 1$, $j = 1, \ldots, k$, where $\mathbf{e}_s$ is a p-dimensional unit vector with $s = \sum_{l=1}^{j-1} k_l + i$, $\mathbf{0}_{p\times(j-1)}$ is a matrix of 0's;

$$\dot{\boldsymbol{\Sigma}}_{b\phi_{ii}} = \boldsymbol{\Lambda}_b\mathbf{e}_i\mathbf{e}_i'\boldsymbol{\Lambda}_b', \quad \dot{\boldsymbol{\Sigma}}_{b\phi_{ij}} = \boldsymbol{\Lambda}_b(\mathbf{e}_i\mathbf{e}_j' + \mathbf{e}_j\mathbf{e}_i')\boldsymbol{\Lambda}_b',\ i,\ j = 1, \ldots, k, \tag{27b}$$

where $\mathbf{e}_i$ and $\mathbf{e}_j$ are of k-dimension; and

$$\dot{\boldsymbol{\Sigma}}_{b\psi_{ii}} = \mathbf{e}_i\mathbf{e}_i',\ i = 1, \ldots, p, \tag{27c}$$

where $\mathbf{e}_i$ is of p-dimension.

Suppose we choose the matrix $\mathbf{T}_b$ as

$$\mathbf{T}_b = (\mathbf{t}_1^{(b)}, \ldots, \mathbf{t}_k^{(b)}, \mathbf{t}_{k+1}^{(b)}, \ldots, \mathbf{t}_{k+p}^{(b)}) = (\boldsymbol{\Lambda}_b \boldsymbol{\Phi}_b^{1/2}, \boldsymbol{\Psi}_b^{1/2}). \tag{28}$$

Then it is obvious that $\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime}$ can be expressed as a linear combination of those in (27b) for $i = 1$ to k, and a linear combination of those in (27c) for $i = k+1$ to $k+p$. So we have

$$\mathrm{vech}(\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime}) \in \mathcal{R}(\dot{\boldsymbol{\sigma}}_b), \ i = 1, \ldots, k+p. \tag{29}$$

According to Lemma 4.1 of Yuan and Bentler [35], (29) implies $\mathbf{V}_b\,\mathrm{vech}(\mathbf{t}_i^{(b)}\mathbf{t}_i^{(b)\prime}) = \mathbf{0}$. Consequently, conditions in (23) are satisfied by model $\boldsymbol{\Sigma}_b(\boldsymbol{\theta})$. Similarly, conditions in (23) are satisfied by model $\boldsymbol{\Sigma}_w(\boldsymbol{\theta})$ if

$$\mathbf{T}_w = (\mathbf{t}_1^{(w)}, \ldots, \mathbf{t}_k^{(w)}, \mathbf{t}_{k+1}^{(w)}, \ldots, \mathbf{t}_{k+p}^{(w)}) = (\boldsymbol{\Lambda}_w \boldsymbol{\Phi}_w^{1/2}, \boldsymbol{\Psi}_w^{1/2}). \tag{30}$$

The statistic T_{LR} will be asymptotically robust if $r_b = r_w = 1.0$ regardless of what the marginal distributions of $\xi_i^{(b)}$ and $\xi_j^{(w)}$ are.

With $\mathbf{T}_b$ as in Eq. (28), the factors and unique variates specified in (6) are independent; the unique variates are also independent. This is the asymptotic robustness condition specified in [1]. As illustrated above, this condition is a special case of the more general conditions given in (23). In Monte-Carlo studies, Eq. (28) is a convenient way for generating nonnormal factors and nonnormal unique variates. Although observed data may be far away from following a normal distribution, the normal theory based LR statistic is asymptotically robust for testing the factor model in conventional covariance structure analysis as well as in the context of multilevel covariance structure analysis. This has been verified by simulations in [37,39].

3. Numerical illustration

For unbalanced data the convergence of $\mathbf{U}_J\boldsymbol{\Gamma}_J$ is too complicated to deal with analytically. We will illustrate its convergence numerically. We will also contrast the results for unbalanced data with those for balanced data. We will use a two-level factor model as specified in (26). With six observed variables and two within- and two between-level factors, the model is generated by

$$\boldsymbol{\Lambda}_b = \begin{pmatrix} 2.0 & 1.5 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.0 & 1.5 & 1.0 \end{pmatrix}', \quad \boldsymbol{\Phi}_b = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix},$$

$$\boldsymbol{\Lambda}_w = \begin{pmatrix} 1.0 & 1.0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 1.0 & 1.0 \end{pmatrix}', \quad \boldsymbol{\Phi}_w = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{pmatrix},$$

and $\boldsymbol{\Psi}_b$ and $\boldsymbol{\Psi}_w$ are diagonal matrices such that all the diagonal elements in $\boldsymbol{\Sigma}_b$ equal 5.0 and all the diagonal elements in $\boldsymbol{\Sigma}_w$ equal 2.0. In parameterizing the model, the factor loadings $\lambda_{31}^{(b)}, \lambda_{62}^{(b)}, \lambda_{31}^{(w)}, \lambda_{62}^{(w)}$ are fixed at 1.0. So there are $q_b = 13$ free parameters in $\boldsymbol{\theta}_b$ and $q_w = 13$ free parameters in $\boldsymbol{\theta}_w$. With $p^* = 21$, there are 16 degrees of freedom in this model.

The matrices $\mathbf{T}_b$ and $\mathbf{T}_w$ are chosen according to (28) and (30), respectively. Three conditions for $\boldsymbol{\xi}$ are chosen. (I) $(\xi_1^{(b)}, \xi_2^{(b)})' \sim N(\mathbf{0}, \mathbf{I}_2)$, $\xi_3^{(b)}$ to $\xi_8^{(b)}$ each follows $(\chi_1^2 - 1)/\sqrt{2}$; $(\xi_1^{(w)}, \xi_2^{(w)})' \sim N(\mathbf{0}, \mathbf{I}_2)$, $\xi_3^{(w)}$ to $\xi_8^{(w)}$ each follows $(\chi_2^2 - 2)/2$. (II) $\xi_1^{(b)}$ and $\xi_2^{(b)}$ each follows

Table 1
CV of the 16 eigenvalues of $\mathbf{U}_J \boldsymbol{\Gamma}_J$ as the average $\bar{n}$ changes

$\bar{n}$	27.5	55	110
(I)	6.11×10^{-5}	1.82×10^{-5}	4.97×10^{-6}
(II)	3.41×10^{-5}	9.72×10^{-6}	2.61×10^{-6}
(III)	1.68×10^{-4}	4.89×10^{-5}	1.33×10^{-5}

$(\chi_3^2 - 3)/\sqrt{6}$, $\xi_3^{(b)}$ to $\xi_8^{(b)}$ each follows $(\chi_2^2 - 2)/2$; $\xi_1^{(w)}$ and $\xi_2^{(w)}$ each follows $(\chi_3^2 - 3)/\sqrt{6}$, $\xi_3^{(w)}$ to $\xi_8^{(w)}$ each follows $(\chi_1^2 - 1)/\sqrt{2}$. (III) $\xi_1^{(b)}$ follows $(\chi_1^2 - 1)/\sqrt{2}$, $\xi_2^{(b)}$ follows $(\chi_2^2 - 2)/2$, $\xi_3^{(b)}$ to $\xi_8^{(b)}$ each follows $[\exp(z_1) - \exp(0.32)]/\sqrt{\exp(0.64)[\exp(0.64) - 1]}$, where $z_1 \sim N(0, 0.8^2)$; $\xi_1^{(w)}$ follows $(\chi_2^2 - 2)/2$, $\xi_2^{(w)}$ follows $(\chi_1^2 - 1)/\sqrt{2}$; $\xi_3^{(w)}$ to $\xi_8^{(w)}$ each follows $[\exp(z_2) - \exp(1/2)]/\sqrt{e(e-1)}$, where $z_2 \sim N(0, 1)$. The third-order moments $E(\xi_i^3) = \zeta_i$ of $N(0, 1)$, $(\chi_3^2 - 3)/\sqrt{6}$, $(\chi_2^2 - 2)/2$, $(\chi_1^2 - 1)/\sqrt{2}$, $[\exp(z_1) - \exp(0.32)]/\sqrt{\exp(0.64)[\exp(0.64) - 1]}$, $[\exp(z_2) - \exp(1/2)]/\sqrt{e(e-1)}$ are respectively 0, 1.633, 2, 2.828, 3.689, 6.185; the fourth-order moments $E(\xi_i^4) = \kappa_i$ of these variables are respectively 3, 7, 9, 15, 34.368, 113.936.

Each of the distribution conditions was evaluated for balanced data with $n = 20, 50, 100$; for unbalanced data with $n_1 = 5k,\ n_2 = 10k, \ldots, n_{10} = 50k$ and $k = 1, 2, 4$. Results for the balanced data apply to any level-2 sample size J. Results for the unbalanced data apply to level-2 sample size $J = 10M$ with $n_1 = \cdots = n_M = 5k$, $n_{M+1} = \cdots = n_{2M} = 10k$, $\ldots$, $n_{9M+1} = \cdots = n_{10M} = 50k$. The average level-1 sample size for the unbalanced data is $\bar{n}_k = k(5 + 10 + \ldots + 50)/10 = 27.5k$.

The sixteen nonzero eigenvalues of $\mathbf{U}_n \boldsymbol{\Gamma}_n$ for balanced data are essentially 1, and departure from 1 cannot be noticed before the tenth decimal place. For unbalanced data, all the nonzero eigenvalues of $\mathbf{U}_J \boldsymbol{\Gamma}_J$ agree with 1 up to at least the third decimal place. The one that differs from 1 most is in condition (III) with $\bar{n}_1 = 27.5$, where the largest eigenvalue is 1.0005. So the convergence is pretty rapid in all the conditions. To get some information of the effect of sample size $\bar{n}$ in different distribution conditions for unbalanced data, Table 1 gives the coefficient of variation (CV) of the sixteen eigenvalues. For each distribution condition, the CV decreases as $\bar{n}$ increases. When $\bar{n}$ changes from 27.5 to 110, the CV decreases more than 10 times in all the conditions. The largest CV is with condition (III) at $\bar{n} = 27.5$. Even for this worst case, with CV $= 1.68 \times 10^{-4}$ and the largest eigenvalue of 1.0005, all the nonzero eigenvalues of $\mathbf{U}_J \boldsymbol{\Gamma}_J$ can be regarded as 1.0 for practical purposes. Actually, one will find that all the nonzero eigenvalues are 1.00 when only reporting up to two decimals and that the CV are 0.000 when only reporting up to three decimals.

4. Discussion

It is well-known that the normal theory LR statistic asymptotically follows a chi-square distribution when data are normal. Because practical data seldom follow the multivariate normality assumption, various efforts have been made to study the robustness of the LR statistic for nonnormal data in the context of linear latent variable models [30,31]. There are nonnormal data options in standard software [4] for conventional SEM. Essentially all statistical programs that provide a multilevel SEM option require the assumption of normal data. It is important to know that results based on such options can also be valid when applied to a nonnormal hierarchical data set.

The asymptotic robustness of T_{LR} obtained here does not necessarily imply that one can trust it for any nonnormal data. Although the data model (6) can generate a variety of nonnormal distributions, it cannot cover all the nonnormal distributions that may be exhibited by practical data. Unfortunately, because the **T** in (6) is not observable, the conditions in (23) are not verifiable. This limitation is parallel to that encountered in robustness research on conventional linear latent variable structures. Furthermore, even when model (6) is true and conditions in (23) are satisfied, one should not expect the LR statistic for such a data set to behave the same as that for a normal data set with the same size. This is because the convergence speed of T_{LR} to a chi-square variate also depends on other distributional factors in addition to the sample sizes. In practice, it is very likely that each level-2 unit $\mathbf{x}_{ij}$, $i = 1, \ldots, n_j$ has a different nonnormal distribution. It is not clear how the LR statistic would behave in such a situation, and further research is needed.

In summary, it is important to know that the LR statistic is asymptotically robust for many nonnormal distributions. Nonetheless, one should not blindly trust it to be so for an arbitrary nonnormal data set in practice.

Acknowledgements

We are thankful to a referee for his/her very constructive comments that have led the paper to an improvement over the previous version.

AppendixA.

This appendix will outline the steps for obtaining $\mathbf{C}_{j12}$ and $\mathbf{C}_{j13}$. It is easy to see

$$\bar{\mathbf{x}}_{.j} = \boldsymbol{\mu} + \mathbf{v}_j + \bar{\mathbf{u}}_{.j} \quad \text{and} \quad n_j \mathbf{S}_j = \sum_{i=1}^{n_j} (\mathbf{u}_{ij} - \bar{\mathbf{u}}_{.j})(\mathbf{u}_{ij} - \bar{\mathbf{u}}_{.j})'.$$

Because $\mathbf{v}_j$ and $\bar{\mathbf{u}}_{ij}$ are independent and $E(\mathbf{v}_j) = E(\mathbf{u}_{ij}) = \mathbf{0}$,

$$\mathbf{C}_{j12} = E(\mathbf{h}_{j1}\mathbf{h}'_{j2}) = E[\mathbf{v}_j \operatorname{vech}'(\mathbf{v}_j\mathbf{v}'_j) + \bar{\mathbf{u}}_{.j} \operatorname{vech}'(\bar{\mathbf{u}}_{.j}\bar{\mathbf{u}}'_{.j})], \tag{A.1}$$

$$\mathbf{C}_{j13} = E(\mathbf{h}_{j1}\mathbf{h}'_{j3}) = E[\bar{\mathbf{u}}_{.j} \sum_{i=1}^{n_j} \operatorname{vech}'(\mathbf{u}_{ij}\mathbf{u}'_{ij}) - n_j\bar{\mathbf{u}}_{.j}\operatorname{vech}'(\bar{\mathbf{u}}_{.j}\bar{\mathbf{u}}'_{.j})]. \tag{A.2}$$

It follows from

$$\bar{\mathbf{u}}_{.j}\,\mathrm{vech}'(\bar{\mathbf{u}}_{.j}\bar{\mathbf{u}}'_{.j}) = \frac{1}{n_j^3}\sum_i\sum_k\sum_l \mathbf{u}_{ij}\,\mathrm{vech}'(\mathbf{u}_{kj}\mathbf{u}'_{lj})$$
$$= \frac{1}{n_j^3}\left\{\sum_{i=k=l} + \sum_{\text{other}}\right\}\mathbf{u}_{ij}\,\mathrm{vech}'(\mathbf{u}_{kj}\mathbf{u}'_{lj})$$

that

$$E[\bar{\mathbf{u}}_{.j}\mathrm{vech}'(\bar{\mathbf{u}}_{.j}\bar{\mathbf{u}}'_{.j})] = \frac{1}{n_j^2}\,E[\mathbf{u}_{ij}\mathrm{vech}'(\mathbf{u}_{ij}\mathbf{u}'_{ij})]. \tag{A.3}$$

It follows from

$$\bar{\mathbf{u}}_{.j}\sum_{i=1}^{n_j}\mathrm{vech}'(\mathbf{u}_{ij}\mathbf{u}'_{ij}) = \frac{1}{n_j}\sum_i\sum_k \mathbf{u}_{ij}\,\mathrm{vech}'(\mathbf{u}_{kj}\mathbf{u}'_{kj})$$
$$= \frac{1}{n_j}\left\{\sum_{i=k} + \sum_{i\neq k}\right\}\mathbf{u}_{ij}\,\mathrm{vech}'(\mathbf{u}_{kj}\mathbf{u}'_{kj})$$

that

$$E[\bar{\mathbf{u}}_{.j}\sum_{i=1}^{n_j}\mathrm{vech}'(\mathbf{u}_{ij}\mathbf{u}'_{ij})] = E[\mathbf{u}_{ij}\mathrm{vech}'(\mathbf{u}_{ij}\mathbf{u}'_{ij})]. \tag{A.4}$$

Note that $\mathbf{v}_j = r_b\mathbf{T}_b\boldsymbol{\xi}_b$. We have

$$\mathrm{vech}(\mathbf{v}_j\mathbf{v}'_j) = r_b^2\mathbf{D}_p^+\mathrm{vec}(\mathbf{T}_b\boldsymbol{\xi}_b\boldsymbol{\xi}'_b\mathbf{T}'_b) = r_b^2\mathbf{D}_p^+(\mathbf{T}_b\otimes\mathbf{T}_b)\mathrm{vec}(\boldsymbol{\xi}_b\boldsymbol{\xi}'_b)$$

and

$$E[\mathbf{v}_j\,\mathrm{vech}'(\mathbf{v}_j\mathbf{v}'_j)] = E(r_b^3)\mathbf{T}_b E[\boldsymbol{\xi}_b\mathrm{vec}'(\boldsymbol{\xi}_b\boldsymbol{\xi}'_b)](\mathbf{T}'_b\otimes\mathbf{T}'_b)\mathbf{D}_p^{+\prime}. \tag{A.5}$$

Similarly,

$$E[\mathbf{u}_{ij}\,\mathrm{vech}'(\mathbf{u}_{ij}\mathbf{u}'_{ij})] = E(r_w^3)\mathbf{T}_w E[\boldsymbol{\xi}_w\mathrm{vec}'(\boldsymbol{\xi}_w\boldsymbol{\xi}'_w)](\mathbf{T}'_w\otimes\mathbf{T}'_w)\mathbf{D}_p^{+\prime}. \tag{A.6}$$

It is easy to see $E[\boldsymbol{\xi}_b\mathrm{vec}'(\boldsymbol{\xi}_b\boldsymbol{\xi}'_b)] = \boldsymbol{\Delta}_b$ and $E[\boldsymbol{\xi}_w\mathrm{vec}'(\boldsymbol{\xi}_w\boldsymbol{\xi}'_w)] = \boldsymbol{\Delta}_w$, where $\boldsymbol{\Delta}_b$ and $\boldsymbol{\Delta}_w$ are given in Lemma 1. The form of $\mathbf{C}_{j12}$ follows from (A.1), (A.3), (A.5) and (A.6). The form of $\mathbf{C}_{j13}$ follows from (A.2)–(A.4) and (A.6).

References

[1] Y. Amemiya, T.W. Anderson, Asymptotic chi-square tests for a large class of factor analysis models, Ann. Statist. 18 (1990) 1453–1463.
[2] T.W. Anderson, Linear latent variable models and covariance structures, J. Econometrics 41 (1989) 91–119.
[3] T.W. Anderson, Y. Amemiya, The asymptotic normal distribution of estimators in factor analysis under general conditions, Ann. Statist. 16 (1988) 759–771.
[4] P.M. Bentler, EQS Structural Equations Program Manual, Multivariate Software, Encino, CA, 1995.
[5] P.M. Bentler, J. Liang, Two-level mean and covariance structures: maximum likelihood via an EM algorithm, in: N. Duan, S. Reise (Eds.), Multilevel Modeling Methodological Advances, Issues, and Applications, Erlbaum, Mahwah, NJ, 2003, pp. 53–70.
[6] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge, 1975.

[7] M.W. Browne, Robustness of statistical inference in factor analysis and related models, Biometrika 74 (1987) 375–384.
[8] M.W. Browne, A. Shapiro, Robustness of normal theory methods in the analysis of linear latent variate models, British J. Math. Statist. Psychol. 41 (1988) 193–208.
[9] S. du Toit, M. du Toit, M. (in press). Multilevel structural equation modeling, in: J. de Leeuw, I. Kreft (Eds.), Handbook of Quantitative Multilevel Analysis, Kluwer, New York.
[10] H. Goldstein, Multilevel Statistical Models, 2nd Edition, Edward Arnold, London, 1995.
[11] H. Goldstein, R.P. McDonald, A general model for the analysis of multilevel data, Psychometrika 53 (1988) 435–467.
[12] J. Hox, Multilevel Analysis: Techniques and Applications, Erlbaum, Mahwah, NJ, 2002.
[13] Y. Kano, Robust statistics for test-of-independence and related structural models, Statist. Probab. Lett. 15 (1992) 21–26.
[14] I. Kreft, J. deLeeuw, Introducing Multilevel Modeling, Sage, London, 1998.
[15] S.-Y. Lee, Multilevel analysis of structural equation models, Biometrika 77 (1990) 763–772.
[16] S.Y. Lee, W.Y. Poon, Analysis of two-level structural equation models via EM type algorithms, Statist. Sinica 8 (1998) 749–766.
[17] J. Liang, P.M. Bentler, An EM algorithm for fitting two-level structural equation models, Psychometrika, 2004, to appear.
[18] N.T. Longford, A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects, Biometrika 74 (1987) 817–827.
[19] N.T. Longford, Regression analysis of multilevel data with measurement error, British J. Math. Statist. Psychol. 46 (1993) 301–311.
[20] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition, Wiley, New York, 1999.
[21] R.P. McDonald, H. Goldstein, Balanced versus unbalanced designs for linear structural relations in two-level data, British J. Math. Statist. Psychol. 42 (1989) 215–232.
[22] A. Mooijaart, P.M. Bentler, Robustness of normal theory statistics in structural equation models, Statist. Neerlandica 45 (1991) 159–171.
[23] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[24] B. Muthén, Multilevel covariance structure analysis, Sociol. Methods Res. 22 (1994) 376–398.
[25] B. Muthén, Latent variable modeling of longitudinal and multilevel data, in: A. Raftery (Ed.), Sociological Methodology 1997, Blackwell Publishers, Boston, 1997, pp. 453–480.
[26] B. Muthén, A. Satorra, Complex sample data in structural equation modeling, in: P.V. Marsden (Ed.), Sociological Methodology 1995, Blackwell Publishers, Cambridge, MA, 1995, pp. 267–316.
[27] W.-Y. Poon, S.-Y. Lee, A distribution free approach for analysis of two-level structural equation model, Comput. Statist. Data Anal. 17 (1994) 265–275.
[28] S.W. Raudenbush, A.S. Bryk, Hierarchical Linear Models, 2nd Edition, Sage, Newbury Park, 2002.
[29] A. Satorra, Asymptotic robust inferences in the analysis of mean and covariance structures, Sociol. Methodol. 22 (1992) 249–278.
[30] A. Satorra, Asymptotic robustness in multiple group linear-latent variable models, Econometric Theory 18 (2002) 297–312.
[31] A. Satorra, Power of χ^2 goodness-of-fit tests in structural equation models: the case of non-normal data, in: H. Yanai, A. Okada, K. Shigemasu, Y. Kano, J.J. Meulman (Eds.), New Developments in Psychometrics, Springer, Tokyo, 2003, pp. 57–68.
[32] A. Satorra, P.M. Bentler, Model conditions for asymptotic robustness in the analysis of linear relations, Computat. Statist. Data Anal. 10 (1990) 235–249.
[33] A. Satorra, P.M. Bentler, Goodness-of-fit test under IV estimation: asymptotic robustness of a NT test statistic, in: R. Gutiérrez, M.J. Valderrama (Eds.), Applied Stochastic Models and Data Analysis, World Scientific, Singapore, 1991, pp. 555–567.
[34] A. Satorra, H. Neudecker, On the asymptotic optimality of alternative minimum-distance estimates in linear latent-variable models, Econometric Theory 10 (1994) 867–883.
[35] K.-H. Yuan, P.M. Bentler, On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions, Statist. Sinica 9 (1999) 831–853.

[36] K.-H. Yuan, P.M. Bentler, Inferences on correlation coefficients in some classes of nonnormal distributions, J. Multivariate Anal. 72 (2000) 230–248.
[37] K.-H. Yuan, P.M. Bentler, On normal theory based inference for multilevel models with distributional violations, Psychometrika 67 (2002) 539–561.
[38] K.-H. Yuan, P.M. Bentler, On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates, Psychometrika 67 (2002) 251–259.
[39] K.-H. Yuan, P.M. Bentler, Eight test statistics for multilevel structural equation models, Comput. Statist. Data Anal. 44 (2003) 89–107.
[40] K.-H. Yuan, P.M. Bentler, On the asymptotic distributions of two statistics for two-level covariance structure models within the class of elliptical distributions, Psychometrika, 2004, to appear.